Eurographics Workshop on Visual Computing for Biology and Medicine (2020)
F. Lekschas and G. Mistelbauer (Poster Chairs)

Towards an Enhanced Interactive Protein Sequence Diagram

Marco Schäfer [1], Arne Cremer[1], Merve Aktürk[1], and Michael Krone [1]

[1]Big Data Visual Analytics in Life Sciences (BDVA), University of Tübingen, Germany

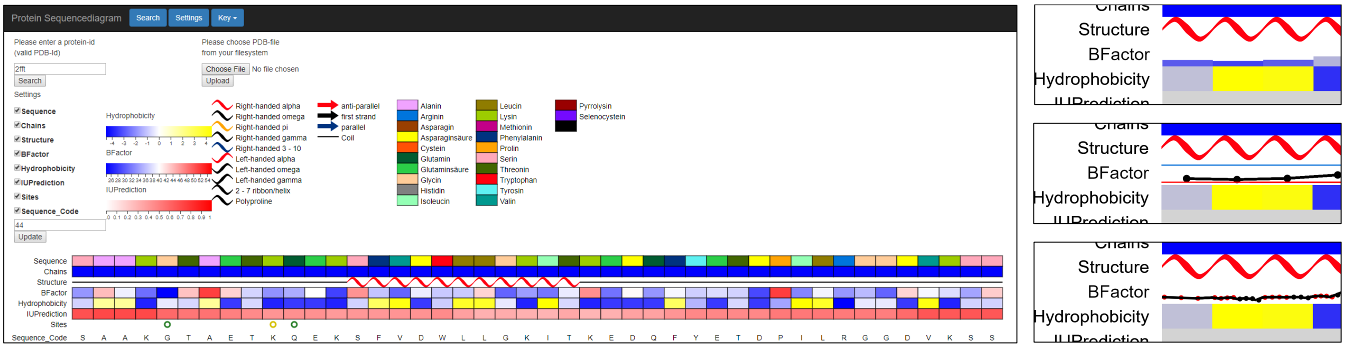

Figure 1: *Left: Our interactive web-based sequence diagram visualization showing a phosphoprotein. Right: Different visual encodings of a numerical value (BFactor). Top: the height of the color-coded rectangles is scaled according to the average BFactor per amino acid, resulting in a bar chart. Middle: line plot, with the minimum and maximum values shown as thin red and blue lines. Bottom: line plot showing the individual values for each atom. The backbone and sidechain atoms of the amino acids are shown as red and black dots.*

Abstract

Sequence diagrams (SD) are a common way to visualize the amino acid sequence of proteins. Usually, not only the sequence is represented, but also other relevant information like binding sites or the secondary structure, if available. Although SD are often used in conjunction with 3D visualizations of the protein structure, they are also useful by themselves for visually analyzing protein data. SD are often visualized as static, precomputed images. Therefore, we present an interactive SD visualization that shows not only the attributes of the protein that are stored in the RCSB Protein Data Bank, but can also visualize additional information provided by external analysis tools. Furthermore, we try to enhance the basic idea of SDs by adding per-atom information instead of showing only per-amino-acid attributes. Figure 1 (right) shows our SD, which was implemented as an interactive, web-based visualization using the JavaScript library D3. Different attributes per amino acid are represented in the rows: the type of amino acid, the chain ID of the amino acid chain, the secondary structure, the BFactor, the hydrophobicity, the predicted intrinsic disorder, and the binding sites. Most of these attributes are encoded using colored rectangles. BFactor, hydrophobicity, and disorder prediction are quantitative attributes for which individual color gradients are used. The secondary structure is graphically depicted via different types of helices and arrows. Binding sites are encoded as colored rings. To provide users with more detailed information, we extended the idea of the classical SD that only shows per-amino-acid attributes. For per-atom attributes like the BFactor, our system computes the overall minimum and maximum values, and the average value per residue. Figure 1 (right) shows three different possible visualizations offered by our extended SD that show either the summary statistics or the individual values per atom. Our SD allows users to interactively hide attribute rows, change the number of amino acid per row, and it offers a tooltip with additional information if the user hovers an amino acid.

Our system is still work in progress, therefore, there are plenty of options for future work: integrating more attributes to support a wider range of analysis tasks, exploring additional visualizations to show more detailed information (e.g., linking to a 3D view), or displaying which residues are bonded (e.g., hydrogen bonds or disulfide bridges) or in contact to residues of another chain. We also want to extend our SD to show multi-sequence alignment, to allow users to visually compare proteins. Furthermore, the user interface and layout of our current prototype also need to be improved. Beside these technical extensions, we also plan to evaluate our proposed per-atom visualization concepts with domain experts, to ensure their utility.

Acknowledgments: *This work was funded by German Research Foundation (DFG) Project ID 391088465 and by Carl-Zeiss-Stiftung.*

CCS Concepts

• ***Applied computing*** → *Computational biology;* • ***Human-centered computing*** → *Visualization systems and tools;*